Middle East Technical University Department of Computer Engineering

CENG 222

Statistical Methods for Computer Engineering

Spring 2019-2020

Homework 4

Last updated on May 27, 2020

Due date: June 20, 2020 23:55PM (GMT+3)

Problem Description

In a merchandise shop, several types of goods are processed daily, where they are eventually distributed to buyers in the form of triplets, in which both every pair has to be compatible with each other and every item is of a distinct type. Equal quantities per goods category arrive in 24 hours time and they are initially stored in the inventory, where their compatibility information is recorded as an undirected graph.

The number of distinct goods types to be processed within a day is a Poisson random variable with $\lambda = 160$.

After the gathering of goods is complete, an undirected graph whose number of vertices equals the number of goods types is constructed such that there exists an edge between two vertices if and only if the goods they indicate are compatible. Otherwise, the said goods types are incompatible, and consequently their reciprocal vertices are not connected by edges.

The undirected graph should be represented as a symmetric square matrix G, whose number of rows (columns) is equal to the number of collected goods types. The entries of the matrix are chosen from the set $\{0, 1\}$, where 1's indicate the existence, and 0's, the absence of corresponding edges. Due to the randomness of the goods characteristics in terms of the compatibility criterion, the undirected graph is formed through Bernoulli trials with p given as the probability that an edge is present, i.e. its associated entry in G is 1. Since we are interested in triplets of distinct goods types given the distribution condition, goods of the same type must be assumed incompatible.

Your task is to estimate the the number different choices the merchant can make for the first shipment of the day, based on given p values. For that purpose, you will compute the number of triangles as sets of three vertices of the graph such that every pair is connected with an edge. You must estimate both the total number of triangles and the ratio of the number of triangles to the number of all distinct triplets of vertices.

Conduct Monte Carlo simulations to answer the subsequent questions. For each option, with probability 0.98, your answer should differ from the true value by no more than 0.03. Use Normal approximation to determine the size of your Monte Carlo simulation. Use `poissrnd` in MATLAB/Octave to sample from the Poisson distribution and Example 5.5 in the textbook to sample from the Bernoulli distribution with the adequate adjustments to generate the symmetric square matrix G of the related random variables.

a. Given $p = 0.012$, indicating that most types of goods are incompatible and G is a sparse graph with very few entries as 1, carry out a Monte Carlo study to estimate the probability that at most 1 distinct choice can be made for the first shipment of the day.

b. Given $p = 0.79$, implying a good chance that any two types of goods in the inventory are compatible, making G a dense graph, perform a Monte Carlo study to estimate the probability that the ratio of the number of possible choices for the first shipment to the the number of different triplets exceeds the threshold 0.5 in one day.

c. For the two highlighted scenarios, estimate both the total number of choices for the first shipments, X, and the rate of potential first shipments to the distinct goods triplets, Y, separately.

d. Estimate $Std(X)$ and $Std(Y)$ under the two experimental settings independently, and comment on the accuracy of the corresponding estimators.

Specifications

To successfuly complete this homework, you do not need to possess a broad background of the graph theory. Yet, concerning this extensively used computer science data structure, you should understand at least the most fundamental issue of graph representation. You may find it useful to check the Wikipedia page on the adjacency matrix representations of unweighted undirected graphs through the link `https://en.wikipedia.org/wiki/Adjacency_matrix` for a quick recap.

The number of triangles, X, in a simple undirected graph with an adjaceny matrix, G, is given by

$$X = \frac{tr(G^3)}{6},$$

where tr is the trace operator. The interested students might want to check the theorem stating that raising the adjacency matrix G to the k-th power gives k-length walks between any two vertices in the underlying graph, and the divisor 6 is included so as to remove the repeated elements in the counting proces. For further information please refer to `https://en.wikipedia.org/wiki/Triangle-free_graph`.

While generating the matrix, you must pay particular attention to make it symmetric. In particular, copy either the upper triangular part to the lower triangular one, or vice versa to complete this task.

Submission

Submit your MATLAB/Octave source code and a short report that describes the Monte Carlo study and answers the questions in parts (a), (b), (c) and (d) via ODTU-Class before the deadline. Your report can be a Word/Latex pdf document or can be a handwritten and scanned pdf/jpeg file. Late submission is allowed with 20 points per day late submission penalty.